PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-231395
(43)Date of publication of application : 22.08.2000

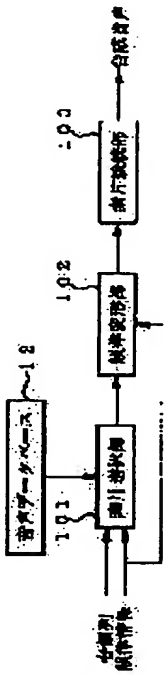
(51)Int.Cl. G10L 13/06
G10L 13/08

(21)Application number : 11-030684
(22)Date of filing : 08.02.1999

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>
(72)Inventor : MIZUNO HIDEYUKI
TANAKA KIMITO
NAKAJIMA SHINYA
ABE MASANOBU

(54) METHOD AND DEVICE FOR SYNTHESIZING VOICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide method and device for synthesizing a voice scalablly changing the voice data and synthetic quality according to a use from a synthesized voice of quality of an extent obtained with a voice synthetic method based on the voice data of low capacity using a voice elemental piece of a length of a phoneme or a syllable etc., until the synthesized voice of high quality similar to a natural voice based on a voice data base of large capacity.
SOLUTION: When a phoneme line and rhythm information are inputted to an elemental piece selection part 101, the elemental piece selection part 101 refers to the inputted phoneme line and rhythm information, and selects the optimum voice elemental piece data from the voice data base 12 to send them to a rhythm deformation part 102. A voice waveform, the rhythm information and rhythm boundary information, etc., are stored in the voice data base 12. The rhythm deformation part 102 deforms the voice elemental piece data selected by the elemental piece selection part 101 so as to be suited to the inputted rhythm information to send them to an elemental piece connection part 103. The elemental piece connection part 103 connects successively the elemental piece data deformed by the rhythm deformation part 102 to generate the synthesized voice.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－231395
（P2000－231395A）
(43)公開日　平成12年８月22日(2000.8.22)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'（参考）
Ｇ１０Ｌ　13/06		Ｇ１０Ｌ　5/04	Ｆ	５Ｄ０４５
13/08		3/00	Ｈ	９Ａ００１

審査請求　未請求　請求項の数６　ＯＬ　（全８頁）

(21)出願番号　特願平11－30684

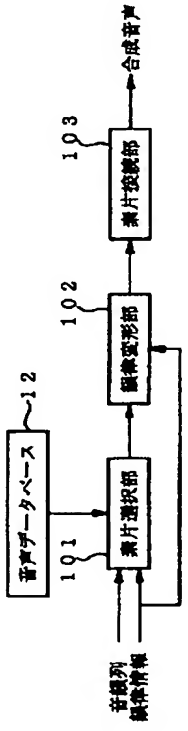
(22)出願日　平成11年２月８日(1999.2.8)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(72)発明者　水野　秀之
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内
(72)発明者　田中　公人
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内
(74)代理人　100064908
弁理士　志賀　正武

最終頁に続く

(54)【発明の名称】　音声合成方法及び装置

(57)【要約】
【課題】　音素または音節等の長さの音声素片を使用した低容量の音声データに基づく音声合成方法で得られる程度の品質の合成音声から、大容量の音声データベースに基づく自然音声と同様の高品質な合成音声まで、用途に応じてスケーラブルに音声データと合成品質を変更可能な音声合成方法及び装置を提供する。
【解決手段】　音韻列と韻律情報が素片選択部１０１に入力されると、素片選択部１０１は入力された音韻列と韻律情報を参照して音声データベース１２より最適な音声素片データを選択して韻律変形部１０２に送る。この音声データベース１２には音声波形，音韻情報，音韻境界情報などが格納されている。韻律変形部１０２は、入力された韻律情報に適合するように素片選択部１０１で選択された音声素片データを変形して素片接続部１０３に送る。素片接続部１０３は韻律変形部１０２で変形された素片データを順に接続して合成音声を生成する。



1

【特許請求の範囲】

【請求項１】　入力された音韻列と韻律情報に対応づけられた音声素片データを音声データベースから選択して順次接続することにより音声信号を合成する音声合成方法において、

前記入力された音韻列を予め決めら　た規則に従って部分音韻列に分解する分解過程と、

対応づけられた音韻列が前記分解された部分音韻列と一致し、かつ該音韻列の前後の音韻が前記部分音韻列の前後の音韻と一致する音声素片データの前記音声データベース中における存在の有無を判断する判断過程と、

前記音声素片データが存在する場合には、当該音声素片データを選択する選択過程と、

前記音声素片データが存在しない場合には、前記分解された部分音韻列を前記入力された音韻列として、前記部分音韻列の長さが予め定められた最小音韻長に分解されるまで前記分解過程と前記判断過程を反復させる過程と、

前記部分音韻列に対応する前記韻律情報を構成する部分韻律情報に応じて、前記選択された音声素片データを韻律変形する過程と、

前記韻律変形を受けた音声素片データを順次接続して音声信号を合成する過程とを有することを特徴とする音声合成方法。

【請求項２】　前記最小音韻長にまで分解された部分音韻列が存在せず、かつ前記最小音韻長が２である場合、前記部分音韻列と該部分音韻列の前後の音韻を含む部分音韻列を連鎖音韻に分解する過程をさらに有し、

前記選択過程では前記連鎖音韻に対応する音声素片データを選択することを特徴とする請求項１記載の音声合成方法。

【請求項３】　一個の部分音韻列について前記選択された音声素片データが複数個存在する場合、それら音声素片データに対応する韻律と前記部分音韻列に対応する韻律との類似性を判断する過程をさらに有し、

前記選択過程では、前記複数個の音声素片データのうち、最も類似性の高い音声素片データを選択することを特徴とする請求項１乃至２記載の音声合成方法。

【請求項４】　入力された音韻列と韻律情報に対応づけられた音声素片データを音声データベースから選択して順次接続することにより音声信号を合成する音声合成装置において、

前記入力された音韻列を予め決められた規則に従って部分音韻列に分解する分解手段と、

対応づけられた音韻列が前記分解された部分音韻列と一致し、かつ該音韻列の前後の音韻が前記部分音韻列の前後の音韻と一致する音声素片データの前記音声データベース中における存在の有無を判断する判断手段と、

前記音声素片データが存在することを条件として、当該音声素片データを選択する選択手段と、

2

前記音声素片データが存在しないことを条件として、前記分解された部分音韻列を前記入力された音韻列として前記分解手段に入力して、前記部分音韻列の長さが予め定められた最小音韻長に分解されるまで前記分解手段と前記判断手段とを反復動作させるように制御する手段と、

前記部分音韻列に対応する前記韻律情報を構成する部分韻律情報に応じて、前記選択された音声素片データを韻律変形する手段と、

前記韻律変形を受けた音声素片データを順次接続して前記音声信号を合成する手段とを具備することを特徴とする音声合成装置。

【請求項５】　前記最小音韻長にまで分解された部分音韻列が存在せず、かつ前記最小音韻長が２であることを条件として、前記部分音韻列と該部分音韻列の前後の音韻を含む部分音韻列を連鎖音韻に分解する手段をさらに具備し、

前記選択手段は前記連鎖音韻に対応する音声素片データを選択することを特徴とする請求項４記載の音声合成装置。

【請求項６】　一個の部分音韻列について前記選択された音声素片データが複数個存在することを条件として、それら音声素片データに対応する韻律と前記部分音韻列に対応する韻律との類似性を判断する手段をさらに具備し、

前記選択手段は、前記複数個の音声素片データのうち、最も類似性の高い音声素片データを選択することを特徴とする請求項４乃至５記載の音声合成装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、テキストを入力しそのテキストに応じた任意の音声を合成する音声合成方法及び装置に関し、特に、主に音韻列と韻律情報とから音声を合成する規則音声合成方法及びこの方法を実現するための装置に関するものである。

【０００２】

【従来の技術】従来の音声合成方法では、あらかじめ、音声素片として音素単位や、ＣＶ，ＶＣＶ，ＣＶＣ（Ｃ：子音，Ｖ：母音）など音韻の調音結合を考慮した単位、３音韻以上のフォルマントを考慮した単位、または前記全ての単位で音声データベースを作成しておき、音声を合成する際に、入力テキストや韻律情報に応じて音声データベース中から適切な素片データを選択して接続することによって音声合成を行っているものが多い（特開昭５９－２０４０９７号公報，特開平１－０７８３００号公報，特開平６－０９５６９２号公報，特開平９－０９０９７２号公報）。この音声合成方法では、合成音声の品質はおおよそ使用する音声データベースの容量と比例しており、容量は少ないが自然音声よりかなり劣ったものから、容量は大きいがある程度高品質なもの

まで様々なものが開発・製品化されている。しかし、それらの製品は全く独立に開発されており互換性等がないため、容量，品質，応答時間などの使用条件に応じて使い分けることが困難である。

【０００３】さらに、近年では大容量な記憶装置の使用コストの低下にともなって、数十分から数時間に及ぶ音声データをそのまま大容量の記憶装置に蓄積し、入力されたテキスト及び韻律情報に応じた適当な基準で大容量の音声データから適当な長さの音声素片を切り出すとともに、入力された韻律情報に従って切り出された音声素片を適切に変形し接続することによって合成する音声合成方法も提案されている（特許第２７６１５５２号）。この方法では大容量の音声データを用意することで、理論的には高品質な合成音声を作成することが可能であるが、大容量の音声データとそれを格納する記憶装置が必要であるためシステム価格が高くなることや、音声データを収集する基準または方法が確立されていないため、必要な品質に見合った最適な規模の音声データを収集することが不可能であること、存在する音声データから適切な音声素片を切り出す最適な規則や方法が確立されていないため、切り出された音声素片が必ずしも適切でなく合成音声全体の品質が安定しないこと等の問題がある。

【０００４】

【発明が解決しようとする課題】この発明は上述した問題点に鑑みてなされたものであり、その目的は、音素または音節等の長さの音声素片を使用した低容量の音声データに基づく音声合成方法によって得られる程度の品質の合成音声から、大容量の音声データベースに基づく自然音声と同様の高品質な合成音声まで、用途に応じてスケーラブルに音声データと合成品質を変更することが可能な音声合成方法及び装置を提供することにある。また、この発明の目的は、大容量の音声データにもとづく音声合成方式の問題を解決し、音声データの収集基準と音声素片の選択規則を明確化することにより、常に適切な音声素片データの選択が保証された高品質な合成音声を実現できる音声合成方法及び装置を提供することにある。

【０００５】

【課題を解決するための手段】以上の課題を解決するために、請求項１記載の発明は、入力された音韻列と韻律情報に対応づけられた音声素片データを音声データベースから選択して順次接続することにより音声信号を合成する音声合成方法において、前記入力された音韻列を予め決められた規則に従って部分音韻列に分解する分解過程と、対応づけられた音韻列が前記分解された部分音韻列と一致し、かつ該音韻列の前後の音韻が前記部分音韻列の前後の音韻と一致する音声素片データの前記音声データベース中における存在の有無を判断する判断過程と、前記音声素片データが存在する場合には、当該音声素片データを選択する選択過程と、前記音声素片データが存在しない場合には、前記分解された部分音韻列を前記入力された音韻列として、前記部分音韻列の長さが予め定められた最小音韻長に分解されるまで前記分解過程と前記判断過程を反復させる過程と、前記部分音韻列に対応する前記韻律情報を構成する部分韻律情報に応じて、前記選択された音声素片データを韻律変形する過程と、前記韻律変形を受けた音声素片データを順次接続して音声信号を合成する過程とを有することを特徴としている。

【０００６】また、請求項２記載の発明は、請求項１記載の発明において、前記最小音韻長にまで分解された部分音韻列が存在せず、かつ前記最小音韻長が２である場合、前記部分音韻列と該部分音韻列の前後の音韻を含む部分音韻列を連鎖音韻に分解する過程をさらに有し、前記選択過程では前記連鎖音韻に対応する音声素片データを選択することを特徴としている。また、請求項３記載の発明は、請求項１乃至２記載の発明において、一個の部分音韻列について前記選択された音声素片データが複数個存在する場合、それら音声素片データに対応する韻律と前記部分音韻列に対応する韻律との類似性を判断する過程をさらに有し、前記選択過程では、前記複数個の音声素片データのうち、最も類似性の高い音声素片データを選択することを特徴としている。

【０００７】また、請求項４記載の発明は、入力された音韻列と韻律情報に対応づけられた音声素片データを音声データベースから選択して順次接続することにより音声信号を合成する音声合成装置において、前記入力された音韻列を予め決められた規則に従って部分音韻列に分解する分解手段と、対応づけられた音韻列が前記分解された部分音韻列と一致し、かつ該音韻列の前後の音韻が前記部分音韻列の前後の音韻と一致する音声素片データの前記音声データベース中における存在の有無を判断する判断手段と、前記音声素片データが存在することを条件として、当該音声素片データを選択する選択手段と、前記音声素片データが存在しないことを条件として、前記分解された部分音韻列を前記入力された音韻列として前記分解手段に入力して、前記部分音韻列の長さが予め定められた最小音韻長に分解されるまで前記分解手段と前記判断手段とを反復動作させるように制御する手段と、前記部分音韻列に対応する前記韻律情報を構成する部分韻律情報に応じて、前記選択された音声素片データを韻律変形する手段と、前記韻律変形を受けた音声素片データを順次接続して前記音声信号を合成する手段とを具備することを特徴としている。

【０００８】また、請求項５記載の発明は、請求項４記載の発明において、前記最小音韻長にまで分解された部分音韻列が存在せず、かつ前記最小音韻長が２であることを条件として、前記部分音韻列と該部分音韻列の前後の音韻を含む部分音韻列を連鎖音韻に分解する手段をさ

らに具備し、前記選択手段は前記連鎖音韻に対応する音声素片データを選択することを特徴としている。また、請求項６記載の発明は、請求項４乃至５記載の発明において、一個の部分音韻列について前記選択された音声素片データが複数個存在することを条件として、それら音声素片データに対応する韻律と前記部分音韻列に対応する韻律との類似性を判断する手段をさらに具備し、前記選択手段は、前記複数個の音声素片データのうち、最も類似性の高い音声素片データを選択することを特徴としている。

【０００９】以上のように、本発明は、音韻情報と韻律情報とから音声を合成する規則音声合成方法及び装置に適用されるものである。そして本発明は、入力された音韻情報に従って音声データベースから音声素片データを選択する際に、音素や音節などの一定の単位での選択または複雑な規則や計算に基づく選択を行うのではなく、音韻情報をある単純な規則に従って部分音韻列に分解し、分解された音韻列およびその前後の音韻環境に適合する音声素片データを音声データベースから選択し、適合する音声素片データが無かった部分音韻列のみをさらに別の単純な規則に従って分解し、その分解された音韻列に適合する音声素片データを音声データベースから選択し、さらに適合する音声素片データが無かった部分音韻列のみ分解し、という多段階の分解と選択を入力音韻列に対応する全ての音声素片データが見つかるまで行うことに特徴を有している。

【００１０】このように多段階の選択を行うことで、最下段の分解規則に対応した最小単位で音声データベースを構成した場合が最も低容量・低品質な用途に対応するとともに、それより上の段階の分解規則に対応した単位の音声素片データを音声データベースに追加することで、より高品質な用途に対応させることが可能となり、また、最上段の分解規則に対応した最長単位の音声素片データが全て音声データベースに存在する場合は最高品質の用途に対応させることが可能となる。そして各段階に対応する音声素片データを音声データベースに追加または削除するだけで、音声合成システムの変更が簡単に実現できる。

【００１１】また、本発明では最終段階の音声素片選択においては、環境を考慮しない連鎖音韻（ＣＶ，ＶＶ，ＶＣ）にもとづく合成方法も適用可能であることに特徴を有している。このようにすることで、環境を考慮して音韻単位で音声素片データを用意した場合は数千～数万個の音声素片データが必要となるのに対し、本発明のように連鎖音韻単位で音声素片データを用意した場合は約千個程度の音声素片データを用意すればよい。そのため少量の記憶装置やメモリなどに音声データベースを格納でき、ＬＳＩ（大規模集積回路）への内蔵用途等にも対応可能となる。

【００１２】また、本発明では各段階で複数の音声素片データが選択された場合、音声素片のピッチパタンが合成すべきピッチパタンともっとも類似する音声素片データを選択することにも特徴を有している。このようにすることで、音声合成時のピッチの変更量を少なくすることができ、合成音声の品質を向上させることが可能となる。また合成すべきピッチパタンと同一のピッチデータをもつ音声素片データを追加することで、ピッチの変形処理が不要になり、その場合の品質は編集音声合成の品質とほぼ同等となる。

【００１３】

【発明の実施の形態】以下、図面を参照してこの発明の一実施形態を述べる。図１に本実施形態による音声合成処理を実現するための音声合成装置の基本構成を示す。図示したように、この音声合成装置は音声データベース１２，素片選択処理を行う素片選択部１０１，韻律変形処理を行う韻律変形部１０２，素片接続処理を行う素片接続部１０３から構成されている。これら各部が行うそれぞれの処理については以下に詳述する。また、図２はこの音声合成処理の手順を示したフローチャートである。最初に図１を参照しながら音声合成装置の全体動作について説明し、その後に、図３及び図４を参照して音声合成装置を構成する幾つかの機能ブロックの構成及びその動作の詳細について説明する。

【００１４】図１に示すように、音韻列と韻律情報が素片選択部１０１に入力される（図２のステップＳ１）と、素片選択部１０１は入力された音韻列と韻律情報を参照して音声データベース１２より最適な波形（素片データ）を選択して韻律変形部１０２に送る（ステップＳ２）。ここで、音声データベース１２には音声波形，音韻情報，音韻境界情報などが格納されているものとする。なお、素片選択部１０１の詳細な構成については後述する。次に、韻律変形部１０２は、入力された韻律情報に適合するように、部分音韻列に対応する韻律情報を構成している部分韻律情報に応じて、前記素片選択部１０１で選択された素片データを変形して素片接続部１０３に送る（ステップＳ３）。

【００１５】ここで、波形の変形方法としては、ＰＳＯＬＡ法（E.Moulines and F.Charpentier, “Pitch-synchro-nous waveform processing techniques for text-to-speech synthesis using diphones”, Speech Communication , Vol.9, pp.453-467, 1990.12）、ＩＰＳＥ法（田中ら，「基本周波数に応じてスペクトル包絡を変形するテキスト合成システム」，信学技報，SP96-130, pp.23-30, 1997.3）、ＳＴＲＡＩＧＨＴ法（河原，「聴覚の情景分析と高品質音声分析変換合成法ＳＴＲＡＩＧＨＴ」，音響学会講演論文集，pp.189-192, 1997.9）などがある。従って、音声データベース１２には、それらの合成方式に応じて最適な形で格納すればよく、必ずしも波形データをそのまま格納する必要はない。例えばＳＴＲＡＩＧＨＴ法を用いるのであれば、事前にＳＴＲＡ

ＩＧＨＴ分析で得られたパラメータを格納しておくことで音声合成時の計算時間が削減できる。そして最後に、素片接続部１０３は、前記韻律変形部１０２で変形された素片データを順に接続し合成音声を生成する（ステップＳ４）。以上が本実施形態による音声合成装置において行われる処理の全体的な流れである。

【００１６】次に、素片選択部１０１における処理の１例について、図３のブロック図と前掲した図２のフローチャートを参照して説明する。なお、図３において図１に示したものと同じ構成要素については同一の符号を付してある。図３に示したように、素片選択部１０１は第一段階分解部２０１，第一段階選択部２０２，韻律マッチング部２０３，第二段階分解部２０４，第二段階選択部２０５，第三段階分解部２０６，第三段階選択部２０７，最終段選択部２０８から構成されている。まず、例えば入力音韻列を" bakuoNga／giNsekaino" （ここで、記号／はアクセント境界を示す）とした場合、第一段階分解部２０１で入力音韻列を例えばアクセント句単位に分　する（ステップＳ２１）。これは、日本語ではアクセント単位でまとまって発声される場合が多く、アクセント句が発声現象の大きなまとまりと考えられるためである。この例では、前記入力音韻列が部分音韻列" bakuoNga" とgiNsekaino" に分解される。

【００１７】次に、第一段階選択部２０２は音声データベース１２から音韻列" bakuoNga"，前音韻環境が語頭（図中の記号＃），後音韻環境が" g" という素片データ、および、音韻列 " giNsekaino"，前音韻環境が" a"，後音韻環境が語尾（図中の記号＃）という素片データを音声データベース１２からそれぞれ検索する（ステップＳ２２）。図で示すとおり、音韻列" giNsekaino" に対応する素片データが見つからず（同ステップが“ＮＯ”）に、音韻列" bakuoNga" に対応する素片データ２１のみ見つかった（同ステップが“ＹＥＳ”）場合、第一段階選択部２０２は素片データ２１のみを韻律マッチング部２０３に送る。この場合、音韻列" bakuoNga" に対応する素片データは１つしかない（ステップＳ２５が“ＮＯ”）ため、韻律マッチング部２０３は素片データ２１をそのまま図１の韻律変形部１０２に送る。また、第一段階選択部２０２は音韻列" giNsekaino" を第二段階分解部２０４に送る（ステップＳ２４）。

【００１８】次に、第二段階分解部２０４は、前記第一段階選択部２０２による音声データベース１２の検索で見つからなかった音韻列" giNsekaino" を例えば音節に母音や撥音の連続を含む単位で分解する（ステップＳ２１）。これは撥音や母音が連続している場合、発声現象的に連続しており音響的にも境界を設定するのが困難であるためである。そしてこの例では、" giN"，“se"，“kai"，“no" の４つの部分音韻列に分解される。次に、第二段階選択部２０５は第一段階選択部２０２と同様に、音声データベース１２から、音韻列 " giN" ，前音韻環境が" a" ，後音韻環境が" s" の素片データと、音韻列" se" ，前音韻環境が" N" ，後音韻環境が" k" の素片データと、音韻列 " kai" ，前音韻環境が" e" ，後音韻環境が" n" の素片データと、音韻列" no" ，前音韻環境が" i" ，後音韻環境が語尾（図中の記号＃）の素片データをそれぞれ検索する（ステップＳ２２）。

【００１９】この結果、図で示すとおり第二段階選択部２０５は " giN" に対応する素片データ２２及び素片データ２３，" se" に対応する素片データ２４，" no" に対応する素片データ２５を韻律マッチング部２０３に送る（ステップＳ２３が“ＹＥＳ”）。この場合、素片データ２４と素片データ２５はいずれも音韻列に対応する素片が１つである（ステップＳ２５が“ＮＯ”）ため、韻律マッチング部２０３は素片データ２１と同様にこれらをそのまま図１の韻律変形部１０２に送る。一方、素片データ２２と素片データ２３（ステップＳ２５が“ＹＥＳ”）については、韻律マッチング部２０３が入力された韻律情報とマッチングを行い、入力韻律情報と最も近い素片データを選択してから図１の韻律変形部１０２に送る（ステップＳ２６）。

【００２０】ここで、韻律の近さの判定方法は使用する音声データベース１２の構成による。例えば、音声データベース１２がピッチのバリエーションについてのみ考慮した音声データベースであれば、入力ピッチパタンと最も近い（最も類似性の高い）ピッチパタンをもつ素片データを選ぶことで十分である。また、特に韻律等を考慮していない音声データベースを使用するのであれば、平均ピッチ，ピッチ形状，時間長，パワーの各韻律パラメータについて、入力された値と素片データの持つ値との差分の絶対値を求め、これら絶対値に対して各韻律パラメータ毎の重み係数を掛けて足し合わせることで韻律コストを求め、その値の小さいものを選ぶことが望ましいと考えられる（広川ら，“波形編集型規則合成法における波形選択関数の検討”，音響学会講演論文集，pp.157-158，1989.3）。この例では、素片データ２２が入力ピッチパタンに近いと判断されたとして、韻律マッチング部２０３は素片データ２２を図１の韻律変形部１０２に送る。

【００２１】次に、第三段階分解部２０６は、前記第二段階選択部２０５による音声データベース１２の検索で見つからなかった（ステップＳ２３が“ＮＯ”，ステップＳ２４）部分音韻列 " kai" を例えば音節に分解する（ステップＳ２１）。これは、音節の構成要素である子音と母音は音響的にも発声現象的にも密接に結びついているため、分離して取り扱うのは音質の劣化を招く可能性が大きいためである。この例では、部分音韻列 " kai" が" ka" と“i" の２つの部分音韻列に分解される。次に、第三段階選択部２０７は第一段階選択部２０２及び第二段階選択部２０５と同様に、音声データベー

ス１２から、音韻列”ka”，前音韻環境が”e”，後音韻環境が”i”の素片データと、音韻列”i”，前音韻環境が”a”，後音韻環境が”n”の素片データをそれぞれ検索する（ステップＳ２２）。

【００２２】図で示すとおり音韻列”i”に対応する素片データ２６が一つ見つかり（ステップＳ２３が“ＹＥＳ”）、音韻列”ka”に対応する素片データが見つからなかった（ステップＳ２３が“ＮＯ”）とする。すると第三段階選択部２０７は、韻律マッチング部２０３に素片データ２６を送り、韻律マッチング部２０３は音韻列”i”に対応する素片が１つだけ（ステップＳ２５が“ＮＯ”）のため、前記同様に素片データ２６を図１の韻律変形部１０２に送る。最後に、最終段選択部２０８は前記第三段階選択部２０７による音声データベース１２の検索で見つからなかった部分音韻列”ka”を選択する（ステップＳ２４，ステップＳ２１～Ｓ２６）。

【００２３】次に、最終段選択部２０８の詳細について図４を参照して以下に説明する。なお、図４において図１又は図３に示したものと同じ構成要素については同一の符号を付してある。この図４には２種類の分解・選択方法の一例について示してある。図４（ａ）では、前記の第一から第三までの分解処理および選択処理と同様に、前記第三段階選択部２０７による音声データベース１２の検索で見つからなかった部分音韻列”ka”を音声の基本単位である音韻に分解する方法の一例について示してある。この例では音韻分解部３０１にて部分音韻列”ka”が”k”と“a”に分解される。

【００２４】次に、音韻選択部３０２では第一段階選択部２０２，第二段階選択部２０５と同様に、音声データベース１２から、音韻”k”，前音韻環境が”e”，後音韻環境が”a”の素片データと，音韻列”a”，前音韻環境が”k”，後音韻環境が”i”の素片データをそれぞれ検索する。図４（ａ）で示すとおり”k”に対応する素片データ３１および”a”に対応する素片データ３２が見つかったとして、音韻選択部３０２は図３の韻律マッチング部２０３に素片データ３１と素片データ３２を送る。韻律マッチング部２０３はこれら各音韻に対応する素片が各々１つのみのため、前記同様にこれら素片データを図１の韻律変形部１０２に送る。なお、もしも対応する素片データが複数あった場合は、前記同様に韻律マッチング部２０３にて入力された韻律情報と最も近い韻律を持つ素片データを選択しそれを韻律変形部１０２に送る。

【００２５】一方、図４（ｂ）では、前記第三段階選択部２０７による音声データベース１２の検索で見つからなかった部分音韻列”ka”を連鎖音韻に分解する方法の一例について示してある。連鎖音韻に分解する理由は、前記図４（ａ）で示す方法と比較した場合に、少ないデータ量であらゆる音韻列の合成が可能となるためである。前記図４（ａ）に基づく方法では数千～数万の素片データが必要なのに対し、図４（ｂ）で示す方法では約１０００個程度の素片データのみでよいため、より少ない記憶容量で音声合成が実現可能となる。この例では、部分音韻列”ka”の前音韻環境が”e”，後音韻環境が”i”であったことから、連鎖音韻分解部３０３はこれを”ek”，“ka”，”ai”に分解して連鎖音韻選択部３０４に送る。

【００２６】次に、連鎖音韻選択部３０４は音声データベース１２から”ek”，”ka”，”ai”である連鎖音韻の素片データをそれぞれ検索する。図４（ｂ）で示すとおり”ek”に対応する素片データ３３，”ka”に対応する素片データ３４，”ai”に対応する素片データ３５が見つかったとして、連鎖音韻選択部３０４は図３の韻律マッチング部２０３に素片データ３３～素片データ３５を送る。この場合、各連鎖音韻に対応する素片データは各々１つのみのため、韻律マッチング部２０３は前記同様にこれら素片データを図１の韻律変形部１０２に送る。なお、もしも対応する素片データが複数ある場合は、前記同様に韻律マッチング部２０３にて入力韻律情報と最も近い韻律を持つ素片データを選択しそれを韻律変形部１０２に送る。またこの後、素片接続部１０３が韻律変形部１０２で変形された素片データ（ステップＳ３）を順に接続して合成音声を生成する（ステップＳ４）が、図４（ｂ）の場合は”k”と“a”が重複するので、このまま素片データを接続するだけでは音韻の重複が避けられない。そのため、素片接続部１０３は素片接続に先立って重複しないように音韻の中間部分をつなぐようにしている。

【００２７】

【発明の効果】以上述べたように、この発明によれば、入力音韻列と入力韻律情報に対して段階的に音声データベースから音声素片データを選択してそれら音声素片データを接続することで出力音声を合成している。このため、低コストの合成システムから高コストではあるが高品質な合成システムまで用途に応じてシステム規模をスケーラブルに変更可能であり、実用性に優れた合成システムの提供が可能である。また、最低段階の選択規則に対応した音声素片データに基づく合成品質は保証されているため、一定以上の品質が保証された合成音声の提供が可能である。

【００２８】また、請求項２又は５記載の発明では、部分音韻列を前後の音韻環境を含めて連鎖音韻に分解して連鎖音韻単位で音声素片データを選択するようにしている。このため、環境を考慮して音韻単位で音声素片データを用意した場合には数千～数万個の音声素片データが必要となるのに対し、連鎖音韻単位で音声素片データを用意することで約千個程度の音声素片データを用意すれば良くなる。そのため、音声データベースを少容量のメモリ等に格納することができ、ＬＳＩに内蔵するなどの用途にも対応することができる。

11　　12

【００２９】また、請求項３又は６記載の発明では、ある部分音韻列について複数の音声素片データが選択された場合に、例えば、音声素片データのピッチパタンが合成すべきピッチパタンに最も類似するものを選択している。このため、例えば音声合成時のピッチの変更量が少なくなって合成音声の品質を向上できるほか、合成すべきピッチパタンと同一のピッチデータを持つ音声素片データを追加すればピッチの変形処理が不要になるため、編集音声合成の品質とほぼ同等の合成音声が得られる。

【図面の簡単な説明】

【図１】　本発明の一実施形態による音声合成装置の構成を示すブロック図である。

【図２】　同実施形態における音声合成方法の手順を示したフローチャートである。

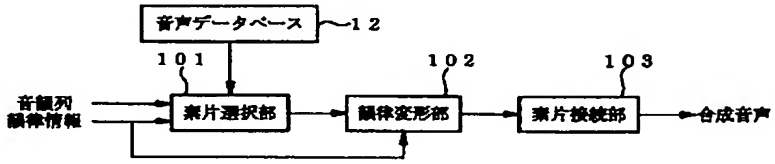
【図３】　図１に示す素片選択部１０１の詳細な構成を示すブロック図である。

【図４】　図３に示す最終段選択部２０８の詳細な構成を示すブロック図であって、（ａ）は部分音韻列を音韻に分解するようにした場合の構成例，（ｂ）は部分音韻列を連鎖音韻に分解するようにした場合の構成例である。

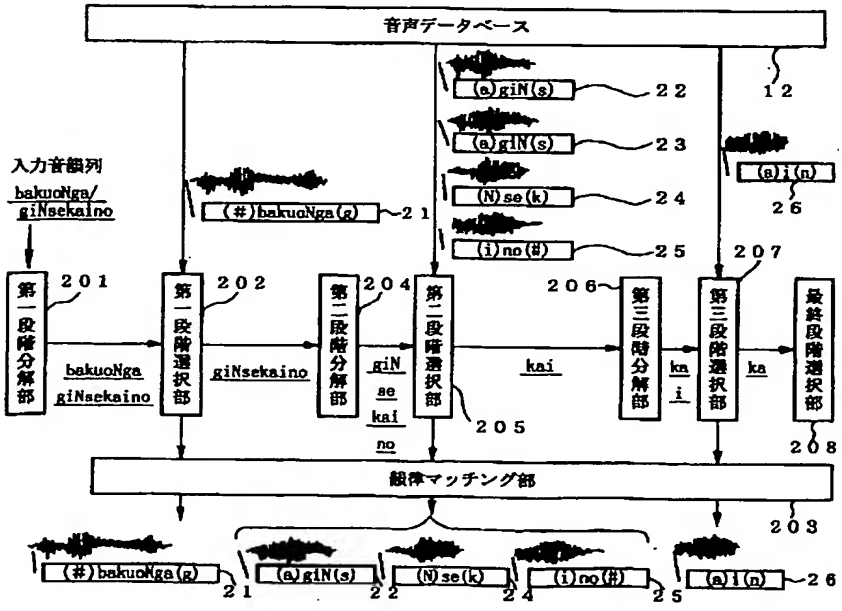
【符号の説明】

１２　音声データベース
２１～２６，３１～３５　素片データ
１０１　素片選択部
１０２　韻律変形部
１０３　素片接続部
２０１　第一段階分解部
２０２　第一段階選択部
２０３　韻律マッチング部
２０４　第二段階分解部
２０５　第二段階選択部
２０６　第三段階分解部
２０７　第三段階選択部
２０８　最終段選択部
３０１　音韻分解部
３０２　音韻選択部
３０３　連鎖音韻分解部
３０４　連鎖音韻選択部

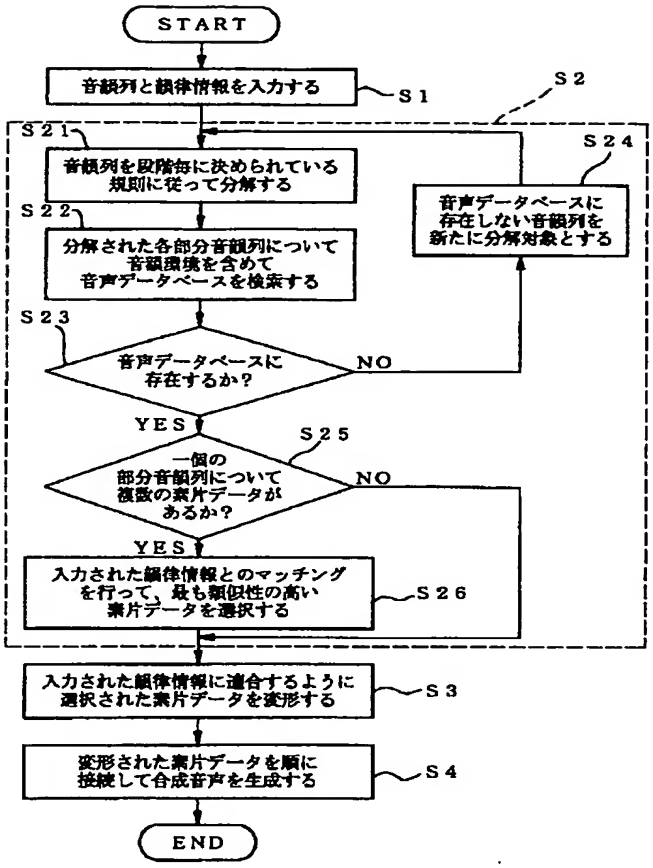
【図１】



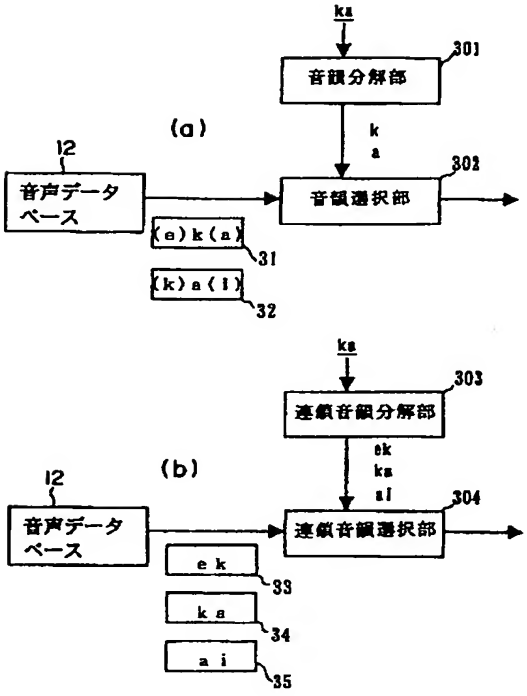
【図３】



【図２】



【図４】



フロントページの続き

(72)発明者　中嶌　信弥
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内

(72)発明者　阿部　匡伸
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内

Ｆターム(参考)　5D045 AA07 AB01
9A001 HH18 JJ01